



EDUCATION AND OCCUPATIONAL OUTCOMES: EVIDENCE FROM BRAZIL

Journal:	<i>International Journal of Manpower</i>
Manuscript ID	IJM-05-2015-0079.R1
Manuscript Type:	Research Paper
Keywords:	Occupation, Education, Dynamics, Development
Note: The following files were submitted by the author for peer review, but marked to be sent in Off-Line.	
Ocupações PNAD Ricardo logs.zip	

SCHOLARONE™
Manuscripts

review

1
2
3 **1. Introduction**
4

5 Education has been seen as a route to prosperity by the governments of many countries,
6 both in the developed and developing world. In the former countries, education has been
7 promoted as a means of securing a comparative advantage in the production of goods
8 and services that embed a high degree of human capital. Many less affluent countries
9 have likewise seen education as a route to development. This raises the question of how
10 successful such policies can be: are countries merely leapfrogging one another in a zero-
11 sum game, or does education offer prosperity for all?
12

13
14 In this paper, we focus on one country that has experienced rapid development in recent
15 times – Brazil. Educational provision, particularly at tertiary level, has expanded greatly
16 over the last two decades. The enhanced skills with which many young people now
17 enter the labour force are likely to impact upon their trajectory through the labour
18 market. In particular, we might expect an increasing proportion of workers to find
19 employment in higher status occupations – typically non-manual jobs in the formal
20 sector. This outcome depends, however, not only on the supply of qualified labour; the
21 demand for such labour equally plays a part in determining outcomes. If supply
22 outstrips demand, then the result of expanding educational provision may be
23 overeducation in the labour market. Demand conditions have evolved rapidly as the
24 impact of globalisation on trading patterns has taken effect. It is therefore hazardous to
25 make inferences about the impact of education policy in any one country from the
26 findings of studies conducted elsewhere, and – regardless of how supply and demand
27 compare – results from a study of this kind are likely to have important implications for
28 policy makers.
29
30

31
32 The hypotheses which we seek to evaluate in this paper relate to these questions: in
33 response to an increase in aggregate educational investment, do individuals’
34 propensities (i) to continue in education and (ii) subsequently to enter higher skill
35 occupations increase?
36

37
38 The conventional approach to studying occupational destination involves estimation of
39 a static discrete choice model. We undertake a static multinomial logit analysis, using
40 cross-section data from the PNAD¹(1993, 1999, 2005, 2013) – a large and rich repeated
41 cross-section data set.² We recognize, however, that this analysis, while instructive, can
42 take us only part of the way towards answering our research question. A fuller answer
43 requires a dynamic analysis capable of examining how an individual’s career path
44 evolves. To examine this, we undertake also a longitudinal analysis using a structural
45 dynamic discrete choice modelling method, with panel data from RAIS-Migra³ (1995 to
46 2006), supplemented with some aggregate data from the Population Census⁴. The
47 application of dynamic analysis using the panel RAIS-Migra data represents an
48 important innovation; most studies of occupational choice in the received literature are
49 based on a static analysis, and the use of relatively newly developed dynamic tools
50
51

52
53 ¹ In Portuguese: Pesquisa Nacional por Amostra de Domicílio.
54 ² We construct separate models for different years in preference to a pooled model with fixed effects for
55 years, because our primary focus is on changes over time in the response of occupational destination to
56 education.
57 ³In Portuguese: Relação Anual de Informações Sociais.
58 ⁴ The Population Census, last conducted in 2010, is assembled by the Instituto Brasileiro de Geografia e
59 Estatística.
60

drawing on panel data is key to understanding how the economic impact of recent changes in education policy and provision will unfold over time.

The paper is structured as follows. We begin with a survey of the relevant literature. We then consider the main methodological approaches used in the study, namely the static approach of multinomial logit analysis and the dynamic discrete choice approach. The latter approach has been applied rather infrequently – it is demanding of data, of programming time, and of real computing time – and so we describe the method in some detail before proceeding to a description of data sources. This is followed by a presentation of the results of our estimation exercises. The paper ends with a discussion and conclusion.

2. Literature Review

Early work in the analysis of occupational choice stems from the seminal contribution of Roy (1951) who focuses on how destination depends upon skills and upon the distribution of returns to skills in each occupation. Workers gravitate to destinations that are most remunerative for them. The distinctive feature of Roy's contribution is in the dynamics. Conditions can change, and workers may respond by changing occupations; how they do so depends on their accumulated characteristics and on the returns that they expect from a large number of possible future career paths. This work thus anticipated later key contributions by Schultz (1961) and Becker (1964) serving to establish that the investment that people make in the development of their own skills, through education and through experience, has a payoff in terms of productivity, and hence in remuneration. People augment their skills by making education and work choices in such a way as to maximize their net return on the investment; the dynamic pattern of such investment, and the way that it manifests itself through job changes, can be complex.

Static and, to a lesser extent, dynamic models of occupational choice have been widely applied to the analysis of occupational choice in developed economies. Variants of the static model have been employed by, *inter alia*, Boskin (1974), Schmidt and Strauss (1975), Ham (1982), Makepeace (1996) and Johnes (1999). In these examples, the emphasis has been on the development of a structural model in which wages explicitly play a key role in determining occupational destination. These analyses use cross-section data on individuals, and thus refer to occupation at a single point in time. The seminal contribution in the area of dynamic modelling is that of Keane and Wolpin (1997), but other important papers include Stinebrickner (2000, 2001a, 2001b), Pavan (2011).

We turn now to consider applications of the static models described earlier to the context of rapidly developing economies – the BRICs. Despite the availability of high quality household data, there have been relatively few analyses of occupational destination in Brazil. In an early study that uses census data, Arriagada and Ziderman (1992) investigate occupational choice, specifically focusing on the extent to which vocational education raises earnings of workers in different occupations. They find that, where there is a good match between the nature of the vocational education and the characteristics of the occupation in which a worker is employed, the benefit attached to education is high, with a Mincerian rate of return of around 22 per cent. This does not

differ significantly from the rate of return to academic education. The estimate compares fairly closely with that of Behrman *et al.* (1996), who find a Mincerian rate of return to education in Brazil of around 19 per cent. These authors undertake further analyses to evaluate the impact of education quality (proxied by teacher experience) on subsequent outcomes, and find that excluding this quality measure from the Mincerian regression leads to upward bias in the estimate of the rate of return.

A particularly important Brazilian study in the present context, albeit one that uses a somewhat different methodology, is that of Ferreira and Leite (2002). These authors conduct an analysis of the impact of educational expansion on the incidence of poverty in the state of Ceará, using data from the 1999 round of the PNAD. Their model involves separate estimation of a number of separate ‘blocks’, each of which explains an aspect of individual behaviour such as occupational choice and education choice. They then use the estimates from these models to simulate the impact of policy change over time. Our model differs from that of Ferreira and Leite in that we model all decisions within a single, dynamic, framework, and consider the impact of policy changes within this framework. But certain aspects of Ferreira and Leite’s work – especially their use of multinomial logit as a means of modelling choice – are nonetheless similar to the approach we take in some of the work reported below.

Further studies of occupational choice in other BRIC countries include approaches based on multinomial logit analyses in India (Khandker, 1992), South Africa (Keswell, 2000) and Russia (Klimova, 2012). Dynamic analyses are rare, but a notable exception is Aggarwal *et al.* (2013) who use a dynamic modelling approach very similar to that which we employ in the present paper. They find that the policy of expanding educational provision leads initially to an increased participation in education and subsequently to an increased propensity for workers to enter non-manual occupations. This exercise, conducted on synthetic panel data from India, thus provides confirmation that education offers similar benefits in the context of a rapidly developing country to those found elsewhere. In Brazil, unlike India, we have the benefit of access to the high quality RAIS-Migra panel data; the present paper therefore represents the first exercise to apply these dynamic methods to a true panel data set in the context of the BRIC countries.

To the best of our knowledge, no dynamic analysis of occupational destination has been conducted in Brazil. So, in this respect, the present paper represents a substantial departure from, and advance upon, the existing literature.

3. Methodology

Empirical implementation of Roy’s ideas about occupational destination required the development of new econometric tools. Under simplifying assumptions – to be specific, assuming that the large state space of possible future trajectories can be reduced and hence be considered as a one-shot problem - a static discrete choice model is sufficient for the analysis. The multinomial logit model, developed by Nerlove and Press (1973), is in many respects the obvious tool for analysing this type of problem.

This model involves the use of maximum likelihood methods to choose the appropriate parameter estimates in the expressions

$$P(Y=j) = \frac{e^{\delta'_j z_i}}{1 + \sum_{k=1}^J e^{\delta'_k z_i}}, j=1,2,\dots,J$$

$$P(Y=0) = \frac{1}{1 + \sum_{k=1}^J e^{\delta'_k z_i}} \quad (1)$$

where the δ terms are parameters and z explanatory variables.

The multinomial logit method is certainly instructive, but some limitations should be noted. The first, well documented in the literature, is that it makes an assumption of the independence of irrelevant alternatives (IIA). That is, it is assumed that the relative odds between two alternative outcomes are unaffected by augmenting the set of possible outcomes. Several partial fixes for this consideration have been suggested in the literature, including nested logit and mixed logit methods.⁵

A second feature of the static multinomial logit analysis that is unappealing in the present context is that it is insufficiently equipped to investigate the impact of policy changes. A change in educational policy in one period is likely to affect labour market outcomes many years into the future, not instantaneously.

In the present paper we therefore report results of a multinomial logit analysis, but we also adopt an additional approach – that of dynamic discrete choice modelling. This approach relaxes the restrictive assumptions needed in modelling occupational choice as a static problem. The dynamic model links theory to empirical application by adopting a set of structural equations in which all possible regime choices are included, and, *at each date, the choice of regime is determined both by instantaneous returns – themselves a function of choices made in previous periods – and by expected returns that depend on choices made in the future.* Given the number of possible trajectories that an individual might take over the periods that remain in his or her lifetime, the state space to be considered in such problems is large, and the analysis of problems of this kind has had to await the development both of sufficiently powerful computing facilities and adequate approximation methods. Since individuals base their decisions in each period on the entirety of this state space – defined across a multiplicity of periods – the problems associated with the IIA assumption are mitigated; adding a possible outcome to the model impacts on choices, but in a way that accounts fully for the dynamic effect of that outcome on returns associated with other choices.

An early attempt to model the dynamic nature of occupational choice in the context of education and work is due to Willis and Rosen (1979). These authors model the decision of when to leave education as an optimal stopping problem. In their model, there is only one post-school outcome, rather than a multiplicity of destinations (including various occupations and life outside the labour force). This type of problem was developed further by Rust (1987) who introduced the nested fixed point algorithm. The extension of this type of model to the case in which, at each point in time, agents make decisions across a multiplicity of options, and where these decisions are

⁵ Soopramanien and Johnes (2001) offer an example of the use of such methods in the context of occupational choice.

conditioned upon decisions made in the past (and determine the nature of options available in the future) is due to Keane and Wolpin (1994, 1997).

The essence of the problem identified by Keane and Wolpin is that in each period, individuals choose between activities. The instantaneous return to each activity depends upon experience which is made up of the schooling and labour market choices that the individual has made in the past. In each period the choice made by the individual therefore impacts on the returns that she can make not only in that period but in every subsequent period. For an individual seeking to maximise her lifetime returns, *the state space is huge, comprising the career paths defined by all possible choices made at each point in time*. Empirical evaluation of such a model requires the adoption of approximation methods. Keane and Wolpin propose the evaluation of expected future returns at a sample of points in the state space, fitting a regression line on the basis of this sample, and using this line (actually a curve – they fit a polynomial) to estimate expected future returns for points outwith the sample. Using these estimates allows estimation of the parameters of the dynamic model in the usual way, using maximum likelihood. Later in this paper we use a variant of the Keane and Wolpin method that allows for regime-specific shocks to be serially correlated.

The literature on human capital investment has, in large measure, concentrated on estimating the rates of return to schooling, considering this schooling as though it were exogenously assigned to individuals in the population. The dynamic modelling approach pursued in the present paper effectively endogenises this choice – making the decision to remain in school dependent on past experience (including shocks) and anticipated future returns – thereby removing sample selection problems.

A feature of the structural modelling approach used here is the close relationship between the theoretical model and the empirical implementation. The analyst begins with an assumed specification of the theoretical model, and estimates the parameters of this model.⁶ While attractive in the sense that this approach involves the estimation of the parameters of the theoretical model itself, not of some loosely defined empirical counterpart, there are some disadvantages. First, a reader might wish to quibble with the precise specification being assumed in the theoretical model; since the empirical implementation is so closely linked to that particular specification, such a quibble assumes empirical importance. Secondly, the close link between theory and estimation means that generic software cannot be developed to estimate models of this kind. In effect, the whole program must be rewritten from scratch each time the specification of the model is subject to a minor modification. These issues have been widely discussed in the literature. Keane (2010), for example, has noted that ‘structural econometric work is just very hard to do’ – and so is not fashionable. We recognise this; we invite the reader therefore to go along with our story while appreciating that no small aspect of the story can be easily tweaked.

In one important respect, however, our task has been easier than that of earlier researchers in this area. A recent survey of structural dynamic discrete choice models by Aguirregabiria and Mira (2010) is accompanied by a website⁷ that offers software that

⁶ This contrasts with more usual practice, which is to develop some theory and then use regression analysis to test whether or not a particular variable influences another in a particular direction consistent with that theory.
⁷http://individual.utoronto.ca/vaguirre/wpapers/program_code_survey_joe_2008.html

has been used by earlier researchers to estimate these models.⁸ The software is written in high level languages (the Keane and Wolpin program, for example, is in fortran), and still requires considerable adaptation before being used to estimate even models that are very similar to those evaluated in the original applications.

4. Data and Empirical Strategy

The two types of analysis conducted in this paper call on the use of two distinct datasets.

4.1 Static Analysis

For the static multinomial logit analysis, we employ the standard large scale Brazilian household survey, namely PNAD. This dataset has been widely used in the literature; see, for example, Arbache *et al.* (2004) and Ribas and Machado (2007). It contains information concerning, *inter alia*, work experience, education and other personal characteristics. We analyse the data at intervals over a period running from the early 1990s through the mid-2010s, this period being coincident with a rapid rise in educational participation. Hence we use data for the years 1993, 1999, 2005 and 2013. Our focus is on young men and women in the 15-35 year age group, the age range over which individuals typically make their choices considering occupational matters. The six possible outcomes for our dependent variable (y) are agriculture ($status=1$), other manual ($status=2$), non-manual ($status=3$), self-employed outside agriculture ($status=4$), in education ($status=5$), and not in work or education ($status=6$)⁹. Throughout our analyses, we weight the data using the expansion factors provided within PNAD.

The number of observations in the various occupation and activity categories differs. According to Table 1, the most common activities are other manual activities, non-manual, and not in work or education. Analysing the changes from 1993 onward, it is clear that both non-manual and education categories increased in importance through the period, the main increase in education being in the 1990s. On the other hand, the

⁸ Another useful recent survey is provided by Keane and Wolpin (2009).

⁹ The education category includes not only those in full-time education, but also those who are working part-time in manual activities while being in education. It is important to highlight that the occupational classifications adopted by the Brazilian Institute of Geography and Statistics (IBGE) in the Brazilian Classification of Occupations (CBO) have changed over the past two decades, in order to be closer to the standards established by the International Labour Organisation, while other occupations have emerged or have lost relevance in the work force. Hence, we developed a means of reconciling these data to a broad categorisation between manual and non-manual. Our reference was the International Standard Classification of Occupations (ISCO-88), using the one digit classification: (1) Legislators, senior officials and managers; (2) Professionals (e.g. physical, mathematical and engineering science professionals); (3) Technicians and associate professionals; (4) Clerks; (5) Service workers, shop and market sales workers; (6) Skilled agricultural and fishery workers; (7) Craft and related trades workers; (8) Plant and machine operators and assemblers; (9) Elementary occupations. The five first categories – from (1) to (5) – are non-manual activities, while the last four categories – from (6) to (9) are manual activities. When the descriptions in our data have a different code in the IBGE 91classification and in the ISCO classification, we adopted the criteria of the most prevalent category. This is the case for the years of 1993 and 1999. For instance, if a particular code is prevalent on manual activities, we will consider all the individuals as manual workers, and vice-versa. Our sample for 2005 has occupations based on the CBO, which follows the ISCO-88.

percentage employed in agriculture declined markedly over the period under investigation; to a lesser extent the same is true of the self-employment category. The proportion of people that are not in work or education also declined over the years either side of the millennium, though this has increased once more over the most recent period.

We summarise the explanatory variables considered in this study and present brief descriptions and basic statistics in Table 2 for 1993, 1999, 2005 and 2013. The explanatory variables can broadly be grouped into personal characteristics (age, age squared, number of years of study, ethnicity dummies – race), family and household characteristics (number of children younger than age 15, number of males working, number of females working, number of males older than 60, number of females older than 60), and regional dummies (27 Brazilian states).

The ethnicity dummies are separated into: (1) aboriginal Brazilians, (2) white, (3) African Brazilians, (4) Asian Brazilians and (5) pardo Brazilians. In Brazil, there are 27 federal units, comprising 26 states and one federal district. The data are also separated by: number of boys and girls in the household; males and females of working age; and males and females older than 60 years of age. The set of explanatory variables also includes household composition: number of boys (aged under 15); number of girls (aged under 15); males of working age; females of working age; males older than 60; females older than 60.

For the years of 1993, 1999, 2005 and 2013, we have 116987, 129364, 149412, and 124745 observations respectively. The average age for our sample is around 24 years. Consistent with the sharp increase during the 1990s in the incidence of being in school, the average number of years of study completed by respondents continues to increase rapidly throughout the period of study, from around 7 in 1993 to more than 10 in 2013. Ethnicity data indicate that about half of all respondents are white, and most others are accounted for by Pardo Brazilians; African Brazilians account for around 5 per cent of all respondents. Over a third of the (weighted) observations are located in one of just two states: São Paulo and Minas Gerais.

4.2 Dynamic Analysis

For the dynamic analysis, we use data from the RAIS-Migra data set over the period 1995-2006.¹⁰ This is a large longitudinal administrative data set which takes the form of an annual census of all formal sector workers. In view of the large size of this data set (the RAIS-Migra has information from 72,386,408 individuals between the years 1995 and 2006), and of the computer intensive nature of the estimation procedure being used, we have taken a random sample of 2509 male workers, all of whom pass through the school leaving age of 14 at some point during the 1995-2006 window.¹¹

¹⁰RAIS-Migra is a data set produced by the Ministry of Labour and Employment (Ministério do Trabalho e Emprego). RAIS (Relação Anual de Informações Sociais) is the annual social information data set; RAIS-Migra refers to an extension of this data set to enable workers to be tracked through time, primarily for the purpose of analysing migration.

¹¹ We used a random number generator to select 5000 workers passing through the age of 14 during the selected time frame. The final sample comprises the males only. Even for this relatively small sample, the computer program took several days (of real time) to run on Lancaster University's high performance cluster.

It should be noted that the RAIS-Migra data provide information only for years in which the worker is employed in the formal sector.¹² A little over a half of all employment in Brazil is in this sector (Hoek, 2007). We can identify the occupation of workers who are in the formal sector, and so we define manual formal sector employment and non-manual formal sector employment as two of the labour market outcomes of interest. Moreover, it is possible to *infer* activity in some other periods from the data that are provided. In particular, we know from RAIS-Migra the individual's highest level of education and so (on the assumption that education is uninterrupted) we know the individual's age when he leaves education. We therefore know that he is in school at all ages younger than this (above the school starting age). Beyond this age, if he is not observed in RAIS-Migra, he must belong to the 'other activity' category (which may include employment in the informal sector or a state of not being employed). In this way, we can construct a complete, balanced, panel of data for our sample, with four categories of labour market outcome: formal sector manual employment, formal sector non-manual employment, education, and other¹³.

Wage data are available for periods when a worker is employed in the formal sector, and these are used in the estimation as a means of identifying the coefficients. Using the consumer price index, these wage data have been deflated to 2005 values. There is a small number of observations where, while the respondent is known to be in formal sector work, wage data are absent. In these cases, the occupation specific average of the real wage is used.

To quantify education policy (educpol)¹⁴ we use the gross enrolment rate in tertiary education for the relevant age group (18-22), calculated from PNAD data (and, for 2000 only – when there was no PNAD – from the Population Census). The series for this variable is reported in Figure 1 and shows a marked increase in the enrolment rate over time. We recognise that any simple measure of this kind has limitations as an indicator of more broadly defined educational policy. An alternative measure might be the share of national income spent on education. Unfortunately, however, the menu from which we can select a measure of educational policy stance is limited, and it is not possible to locate a consistent series for the whole time period under investigation for any expenditure based measures. The availability of microdata collected from individuals allows the use of a participation based measure, and so this is what we employ in our analysis. Furthermore, we recognise that each 18-22 year old who continues in education adds to the enrolment rate, but note that, given the high number of individuals, this should not result in endogeneity problems.

¹² The formal labour market is subject to a plethora of regulations, in particular covering issues such as severance payments and the requirement to provide notice of termination of contract.

¹³ Since RAIS-Migra refers to formal sector workers, those who do not *at any time* work in this sector are not captured in our data set. This means that there our results may be subject to selection bias for which, given the nature of the data set, we are unable to control.

¹⁴ The choice of education policy variable (educpol) presented something of a challenge in that consistent time series for many of the conventional measures (such as public expenditure on education as a percentage of GDP) are not readily available for all years in our study. Commonly used sources of data such as the World Development Indicators (World Bank, 2006) have gaps for certain years.

5. Empirical Results

In this section we report, first, the results of the static multinomial logit modelling exercise. Results obtained using the dynamic discrete choice model follow.

5.1 Multinomial logit models

We consider six labour market outcomes: (i) agricultural work; (ii) other manual employment; (iii) non-manual employment; (iv) self-employment outside agriculture; (v) in education; and (vi) not in work or education. Explanatory variables are: years of schooling, age, age squared, ethnicity, a full set of region dummies, and a variety of household composition variables. The latter comprise counts of: working age males; working age females; males older than 60; females older than 60; boys; and girls in the household. The standard regions are: Rondônia; Acre; Amazonas; Roraima; Pará; Amapá; Tocantins; Maranhão; Piauí; Ceará; Rio Grande do Norte; Paraíba; Pernambuco; Alagoas; Sergipe; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro; Paraná; Santa Catarina; Rio Grande do Sul; Mato Grosso do Sul; Mato Grosso; Goiás; and Distrito Federal.

In common with Ferreira and Leite (2002), we model occupational choice as a reduced form, choosing not to include an earnings variable as a determinant of choice, but rather including characteristics typical of those found in Mincerian earnings functions as measures of earnings potential. The adoption of a reduced form approach allows us to finesse issues of endogeneity and sample selection bias.

In Tables 3, 4 and 5, we report the marginal effects on the years of schooling variable, separately for each year, and separately for males, females, and all respondents. In general, the results are robust and highly significant. It is clear that in all years, schooling raises the probability with which an individual enters non-manual work, and reduces the probability with which an individual enters manual work. Schooling also raises the probability of continuing in education. For women, in the earlier years, schooling raises the probability of entering self-employment outside of agriculture (but decreases it, albeit insignificantly so, in 2005). This could conceivably reflect gender discrimination; if highly educated women find that their opportunities as employees are limited, they may decide to set up their own businesses.

The estimated coefficients on the control variables are not of primary interest in the context of the present paper, but are nonetheless worthy of some comment. Over the age range under consideration, marginal increases in age reduce the propensity to be ‘not in work or education’ and also reduce the propensity to be in education. The propensity to be in manual employment initially rises with age, but beyond about 29 years of age it then falls. The household composition variables paint a complex picture, but it is notable that higher numbers of children in a household reduce the propensity for the respondent to be ‘not in work or education’. Coefficients on ethnicity dummies are typically significant, with the African Brazilians category obtaining worse outcomes (in the sense that they are more likely to be ‘not in work or education’) than other groups.

Finally the state dummies reveal substantial regional disparities, with outcomes being markedly better in Mato Grosso than elsewhere.¹⁵

We have conducted an extensive series of robustness checks on these results – including estimation over a different age group (23-35), alternative specifications of the schooling variable (including, in separate estimations, a quadratic term in years of schooling and dummies for highest level attained), and pooled estimation; we have also confirmed that we have an appropriate disaggregation of regimes using the test due to Cramer and Ridder (1991). We do not report these results of all these checks here for reasons of space but we note that the qualitative findings are unaffected by these changes in specification.

It is clear from the above that *an increased incidence of education raises the probability with which individuals remain in education, and the probability with which they enter employment as non-manual workers*. The marginal effects – around 0.01 and 0.04 respectively – are stable across genders and over time. It is clear therefore that national investment in education has a direct impact on occupational outcomes, leading to more workers entering non-manual jobs. The analysis does not suggest either a strengthening or a weakening of this effect over recent years. Consequently, in common with analyses conducted in developed economies, it appears that, for Brazil, growth of the demand for highly qualified labour is at least keeping pace with increases in supply. We investigate this further as we turn to consider the dynamic modelling of destination below.

5.2 Dynamic discrete choice model

In this section we evaluate the dynamic model, taking seriously the starting point provided by Keane and Wolpin. We thus begin with the following instantaneous reward functions:

$$\begin{aligned} R_{1t} &= w_1 = \alpha_{10} + \alpha_{11}S_t + \alpha_{12}X_{1t} + \alpha_{13}X_{2t} + \varepsilon_{1t} \\ R_{2t} &= w_2 = \alpha_{20} + \alpha_{21}S_t + \alpha_{22}X_{1t} + \alpha_{23}X_{2t} + \varepsilon_{2t} \\ R_{3t} &= \beta_0 + \beta_1 I(s_t \geq 12) + \beta_2 \text{educpol} + \varepsilon_{3t} \\ R_{4t} &= \gamma_0 + \varepsilon_{4t} \end{aligned} \quad (2)$$

Here s refers to years of schooling received prior to the current period t , x_1 is years of experience in occupation 1, and x_2 is years of experience in occupation 2. The terms R_1 through R_4 denote respectively the instantaneous returns to working in occupation 1 (non-manual occupations in the formal sector), occupation 2 (manual occupations in the formal sector), or schooling, or other activity (which may include other work, unemployment, or absence from the labour force). In the case of the first two outcomes, we observe the wages, w_1 and w_2 respectively, and these are incorporated into the modelling procedure. The ε terms represent alternative-specific, period-specific, random shocks. These are crucial in determining why some workers take certain paths through their career while others take others. The first term in the instantaneous reward for schooling equation indicates that we expect the one-period ‘reward’ associated with schooling at tertiary level, β_1 , to be negative owing to the payment of tuition fees. The second term in that equation is intended to capture the effect of education policy

¹⁵ For reasons of space, we do not report the full results here, but they are available from the authors on request.

(educpol) on the decision to stay on at school, and the sign and magnitude of the coefficient attached to that variable, β_2 , is therefore of primary interest in the present study.

The elements of the state space change over time following the rules

$$\begin{aligned}x_{1t+1} &= x_{1t} + d_{1t} \\x_{2t+1} &= x_{2t} + d_{2t} \\s_{t+1} &= s_t + d_{3t}\end{aligned}\tag{3}$$

where d_k is the binary decision indicator attached to the k th regime. Consider the individual's choice in the final period, T . He must, at time $T-1$, choose regime according to

$$E_{\max}\{R_{1T}, R_{2T}, R_{3T}, R_{4T} | S_{T-1}, d_{kT-1}\}\tag{4}$$

where S denotes the set of choices that have been made heretofore. This expectation can be calculated as a multivariate integral, comprising the returns to each regime in each period, information about the state space and about shocks. For each period prior to $T-1$, a similar expectation must be maximised, the individual thus solving a complex optimisation problem over the state space of all possible career paths. It is the size of this state space that necessitates use of an approximation method such as that devised by Keane and Wolpin and used here.

As with any such method, a number of parameters need to be set by the analyst in order to proceed. For the simulation used to evaluate the regime that yields the greatest expected future return, we use 500 draws; we evaluate the expected return at 300 randomly chosen points in the state space and use the interpolation method for all other points. The discount parameter is set at 0.95. The convergence toward the maximum likelihood solution is deemed to be complete when further iterations fail to achieve an improvement in the log likelihood that exceeds 0.001%. Where not specified above, the values of all parameters used in the estimation program are those set as defaults by Keane and Wolpin.

Parameter estimates are reported in Table 6, and are broadly in line with our prior expectations. The key finding is that educpol raises the propensity of respondents to stay in education. Moreover, educational attainment subsequently increases the propensity to be in formal sector work relative to other destinations – though surprisingly it has a greater effect on entry to manual as opposed to non-manual work in the formal sector. This may reflect aspects of the development path in Brazil, where employment in the production industries continued to grow rapidly over the period under consideration. It may also reflect a lack of employment opportunities in higher status jobs. This finding may usefully be contrasted with the situation in India (Aggarwal *et al.*, 2013) where, with service sector led growth, highly educated workers are predominantly absorbed in the non-manual sector.

It is important to interrogate this finding further, not least because it contrasts with the results of the static analysis reported earlier. Note that the estimated coefficient α_{13} is positive, suggesting that experience in manual formal sector work serves to boost the income that a worker can expect to earn in non-manual formal sector work. Meanwhile

α_{23} is negative. It seems therefore that education serves to improve routes into the formal sector, and that these routes might often involve passing through a period of employment in manual work before progressing to non-manual roles. This finding illustrates a major benefit of the dynamic approach; while the static multinomial logit model predicts that education raises the probability of finding formal sector non-manual employment, the dynamic model shows that this happens through quite a complex career path involving a spell, between education and non-manual employment, in manual work.

The high value of the ρ_{33} parameter is suggestive of unobserved heterogeneity across individuals, and this impacts upon the returns that are available to education. This heterogeneity could, in principle, be modelled by separately evaluating coefficients for respondents that come from different family backgrounds. There is, however, little relevant information on this in RAIS-Migra and so this exercise must be left for a future time when more data are available.

Following Keane and Wolpin (1994, 1997) we evaluate standard errors using the outer product of numerical first derivatives. Keane and Wolpin note that there may be a downward bias associated with these standard errors. Perhaps owing to this, the (very) high t statistics reported in Table 6 for most of the coefficients seem to be quite typical for this type of model. Moreover, we note that the *educpol* variable is clustered across all observations in a given year. We are not aware of any literature that allows correction for such clustering in this context, but note that this too will likely bias the standard error downwards. Our estimates of the standard errors should therefore be interpreted with some measure of caution.

That said, the broad picture that emerges both from the static analysis using PNAD data and from the dynamic analysis using data from RAIS_Migra is consistent. The static analysis reveals that the marginal effect of years of education on the 'not in work or education' outcome declines over time over the period spanning the turn of the millennium, suggesting that the increase in educational enrolment is coincident with improved outcomes. Changes over time in the impact of education on the type of employment gained are ambiguous, however; the propensity for this impact to result in employment in non-manual jobs falls between 1993 and 1999 before rising over the next period, and the propensity to enter manual employment increases over this time frame. These results tie in with those of the dynamic analysis that suggest that (since $\alpha_{21} > \alpha_{11}$) returns to education are initially observed to be highest for manual work, but that (since $\alpha_{13} > \alpha_{23}$) workers with experience in manual occupations are well placed subsequently to receive high returns in non-manual jobs.

For future research, we note that the 'not in schooling or formal sector employment' category derived from the RAIS-Migra data is broad; were it possible to disaggregate this category, we might conceivably find that expanding educational opportunity results in substantially lower levels of non-employment. A number of studies that employ dynamic discrete choice methods divide the population into subgroups (based, for example, on family income). Data limitations have not allowed us to do this, and it may be the case that a more refined specification of the model could identify stronger policy effects. Finally, it should be remembered that the way in which the RAIS-Migra data are collected inevitably result in some selection bias, since only workers that are *at some stage* employed in the formal sector are included in the data.

6. Discussion and Conclusions

Brazil’s economy has undergone substantial change in recent years. Despite the recent downturn, the country is often identified as a prospective engine of global growth. With such speedy development comes a change in the industry mix and a consequent change in the demand for skills. In common with other BRICs, Brazil’s response has been to increase participation in post-compulsory education; as we have seen, participation in tertiary education doubled over the course of a decade. A question then arises of how well matched are the rapid increases in the demand for and the supply of highly qualified labour. If supply outstrips demand, then education might – in contrast to the experience of other countries – become a less attractive route to success in the labour market than in the past. This in turn might provoke a policy response aimed at promoting the growth of knowledge driven sectors (Rodrik, 2004).

The results reported above throw light on this mechanism in the Brazilian context. A policy of education expansion increases the propensity of individuals to stay longer in education (unsurprisingly), and – critically from the point of view of the main hypothesis of this paper – later increases also the likelihood with which workers find jobs in the formal sector; these are positive, and policy-relevant, outcomes. More particularly, the education expansion increases the likelihood of workers selecting into manual formal sector jobs – a result that we did not expect, but which may reflect peculiarities of the current state of development of the Brazilian economy – before possibly later transiting into non-manual employment.

Our results should be treated with a measure of caution. In particular, it should be noted that the approach taken in the dynamic modelling assumes that, at the outset of their working lives, individuals differ only in the random shocks that they encounter. It may well be the case that, with access to better data, different types of individual could be identified, and that improvements to the model fit could be secured by modelling these types in a distinct fashion. Moreover, our findings may be sensitive to the particular time frame over which the analysis has been conducted. As the Brazilian economy evolves, it is possible that the balance of occupations which highly educated workers enter will change. Given the speed of change in this particular economy, the call for further research with which many papers end seems particularly apposite here.

5. References

- Aggarwal, Aradhna, Geraint Johnes, Ricardo Freguglia and Gisele Spricigo (2013) Education and labour market outcomes; evidence from India, *Indian Journal of Labour Economics*, 56, 331-347.
- Aguirregabiria, Victor and Pedro Mira (2010) Dynamic discrete choice structural models: a survey, *Journal of Econometrics*, 156, 38-67.
- Arbache, Jorge Saba, Andy Dickerson and Francis Green (2004) Trade liberalisation and wages in developing countries, *Economic Journal*, 114, F73-F96.
- Arriagada, Ana-Maria and Adrian Ziderman (1992) Vocational secondary schooling, occupational choice and earnings in Brazil, World Bank Working Paper Series 1037, http://siteresources.worldbank.org/BRAZILINPOREXTN/Resources/3817166-1185895645304/4044168-1186326902607/45pub_br65.pdf
- Audretsch, David B., Werner Bönte and JagannadhaPawan Tamvada (2007) Religion and entrepreneurship, *Centre for Economic Policy Research Discussion Papers number* 6378, London.
- Becker, Gary S. (1964) *Human Capital*, Chicago: University of Chicago Press.
- Behrman, Jere, Nancy Birdsall and Robert Kaplan (1996) The quality of schooling and labor market outcomes, ch.8 of Nancy Birdsall and Richard Sabot (eds) *Opportunity Forgone: Education in Brazil*, Baltimore: Johns Hopkins University Press.
- Boskin, Michael (1974) A conditional logit model of occupational choice, *Journal of Political Economy*, 82, 389-398.
- Cramer, Jan S. and Geert Ridder (1991) Pooling states in the multinomial logit, *Journal of Econometrics*, 47, 267-272.
- Ferreira, Francisco H.G. and Phillippe George Leite (2004) Educational expansion and income distribution: a microsimulation for Ceara, in Shorrocks, Anthony and van der Hoeven, Rolph (eds) *Growth, inequality and poverty: prospects for pro-poor economic development*, Oxford: Oxford University Press, <http://www.econ.puc-rio.br/pdf/td456.pdf>
- Ham, John C. (1982) Estimation of a labour supply model with censoring due to unemployment and employment, *Review of Economic Studies*, 49, 335-354.
- Hoek, Jasper (2007) Labor flows in formal and informal markets in Brazil, IZA/World Bank Conference on Employment and Development, Bonn, 8 June, http://www.iza.org/conference_files/worldb2007/hoek_j1567.pdf
- Johnes, Geraint (1999) Schooling, fertility and the labour market experience of married women, *Applied Economics*, 31, 585-592.

Keane, Michael P. (2010) Structural v atheoretic approaches to econometrics, *Journal of Econometrics*, 156, 3-20.

Keane, Michael P. and Kenneth I. Wolpin (1994) The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence, *Review of Economics and Statistics*, 76, 648-672.

Keane, Michael P. and Kenneth I. Wolpin (1997) The career decisions of young men, *Journal of Political Economy*, 105, 473-522.

Keane, Michael P. and Kenneth I. Wolpin (2009) Empirical application of discrete choice dynamic programming models, *Review of Economic Dynamics*, 12, 1-22.

Keswell, Malcolm (2000) Labour market dynamics in South Africa: evidence from KwaZulu-Natal Province, available at <http://www.tips.org.za/files/397.pdf>.

Khandker, Shahidur R. (1992) Earnings, occupational choice, and mobility in segmented labor markets of India, World Bank discussion paper 154, available at http://www-wds.worldbank.org/external/default/WDSPContentServer/WDSP/IB/1999/10/23/000178830_98101903550855/Rendered/PDF/multi_page.pdf.

Klimova, Anastasia (2012) Gender differences in determinants of occupational choice in Russia, *International Journal of Social Economics*, 39, 648-670.

Makepeace, Gerald H. (1996) Lifetime earnings and the training decisions of young men in Britain, *Applied Economics*, 28, 725-735.

Nerlove, Marc and S. James Press (1973) Univariate and multivariate log-linear and logistic models, R-1306-EDA/NIH, Santa Monica: Rand, available at <http://www.rand.org/pubs/reports/2006/R1306.pdf>.

Pavan, Ronni (2011) Career choice and wage growth, *Journal of Labor Economics*, 29, 549-587.

Ribas, Rafael Perez and Ana Flavia Machado (2007) Distinguishing chronic poverty from transient poverty in Brazil: developing a model for pseudo-panel data, http://nip-lac.org/uploads/Rafael_Perez_Ribas.pdf

Rodrik, Dani (2004) Industrial policy for the twenty-first century, available at <http://www.hks.harvard.edu/fs/drodrik/Research%20papers/UNIDOSep.pdf>.

Roy, Andrew Donald (1951) Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3, 135-146.

Rust, John (1987) Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher, *Econometrica*, 55, 999-1035.

Schmidt, Peter and Robert P. Strauss (1975) The prediction of occupation using multiple logit models, *International Economic Review*, 16, 471-486.

Schultz, Theodore W. (1961) Investment in human capital, *American Economic Review*, 51, 1-17.

Soopramanien, Didier and Geraint Johnes (2001) A new look at gender effects in participation and occupation choice, *Labour*, 15, 415-443.

Stinebrickner, Todd R. (2000) Serially correlated variables in dynamic discrete choice models, *Journal of Applied Econometrics*, 15, 595-624.

Stinebrickner, Todd R. (2001a) Compensation policies and teacher decisions, *International Economic Review*, 42, 751-779.

Stinebrickner, Todd R. (2001b) A dynamic model of teacher labor supply, *Journal of Labor Economics*, 19, 196-230.

Willis, Robert J. and Sherwin Rosen (1979) Education and self-selection, *Journal of Political Economy*, 87, S7-S36.

World Bank (2006) World development indicators, 2006, World Bank: Washington DC.

Table 1: Occupation Categories (%)

Y	1993	1999	2005	2013
Agriculture	12.52	10.16	9.22	5.44
Other manual	12.36	11.32	12.37	12.91
Non-manual	30.32	29.86	32.98	36.16
Self-employed outside agriculture	10.57	10.20	9.25	8.34
In education	11.67	16.87	16.27	16.69
Not in work or education	22.56	21.59	19.91	20.46

Source: PNAD.

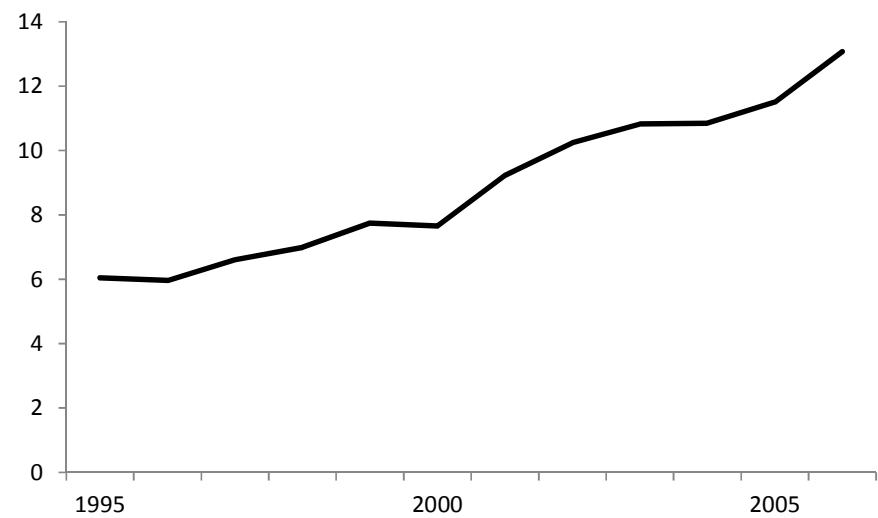
Table 2: Descriptive Statistics

variable	1993		1999		2005		2013	
	mean	sd	mean	sd	mean	sd	mean	sd
Age	24.30	6.09	24.21	6.14	24.40	5.98	24.88	6.14
Schooling	7.24	3.91	8.10	3.87	9.19	3.76	10.41	3.51
Aboriginal	0.001	0.032	0.002	0.041	0.002	0.044	0.003	0.058
White	0.538	0.499	0.530	0.499	0.477	0.499	0.438	0.496
African	0.051	0.221	0.054	0.225	0.067	0.249	0.084	0.278
Asian	0.004	0.065	0.004	0.062	0.004	0.066	0.004	0.066
Pardo	0.405	0.491	0.411	0.492	0.450	0.498	0.470	0.499
Rondônia	0.006	0.074	0.006	0.076	0.009	0.092	0.009	0.095
Acre	0.002	0.045	0.002	0.049	0.004	0.060	0.004	0.066
Amazonas	0.012	0.107	0.013	0.113	0.019	0.137	0.021	0.143
Roraima	0.001	0.034	0.001	0.037	0.002	0.049	0.003	0.051
Pará	0.019	0.137	0.021	0.143	0.041	0.198	0.042	0.202
Amapá	0.002	0.043	0.003	0.051	0.004	0.061	0.004	0.064
Tocantins	0.006	0.080	0.007	0.083	0.007	0.086	0.007	0.086
Maranhão	0.031	0.174	0.032	0.177	0.036	0.186	0.035	0.184
Piauí	0.018	0.132	0.017	0.129	0.017	0.129	0.016	0.125
Ceará	0.042	0.200	0.043	0.203	0.045	0.206	0.046	0.209
Rio Grande do Norte	0.017	0.129	0.017	0.129	0.017	0.129	0.017	0.131
Paraíba	0.021	0.145	0.021	0.143	0.020	0.142	0.020	0.140
Pernambuco	0.049	0.215	0.048	0.213	0.047	0.212	0.046	0.211
Alagoas	0.017	0.131	0.017	0.130	0.017	0.128	0.017	0.128
Sergipe	0.011	0.102	0.011	0.105	0.011	0.106	0.012	0.107
Bahia	0.079	0.270	0.082	0.275	0.078	0.269	0.076	0.264
Minas Gerais	0.110	0.313	0.107	0.309	0.102	0.302	0.101	0.302
Espirito Santo	0.019	0.138	0.019	0.137	0.019	0.136	0.019	0.138
Rio de Janeiro	0.087	0.282	0.081	0.272	0.076	0.265	0.077	0.266
São Paulo	0.225	0.418	0.226	0.419	0.217	0.412	0.213	0.410
Paraná	0.061	0.239	0.059	0.236	0.053	0.225	0.054	0.226
Santa Catarina	0.033	0.178	0.032	0.177	0.032	0.175	0.032	0.176
Rio Grande do Sul	0.061	0.238	0.058	0.234	0.052	0.222	0.049	0.217
Mato Grosso do Sul	0.013	0.113	0.013	0.113	0.013	0.113	0.013	0.113
Mato Grosso	0.016	0.125	0.016	0.126	0.016	0.127	0.017	0.128

Goiás	0.030	0.170	0.032	0.175	0.032	0.176	0.033	0.179
Distrito Federal	0.013	0.113	0.015	0.120	0.014	0.119	0.015	0.122
# boys in household	0.70	0.91	0.59	0.83	0.51	0.77	0.44	0.70
# girls in household	0.70	0.91	0.58	0.82	0.49	0.76	0.42	0.69
# working age men in household	1.47	1.03	1.45	0.99	1.39	0.95	1.32	0.88
# working age women in household	1.45	0.88	1.43	0.85	1.38	0.80	1.35	0.76
# old men in household	0.08	0.27	0.07	0.26	0.07	0.26	0.07	0.26
# old women in household	0.08	0.27	0.08	0.27	0.08	0.27	0.09	0.28

Source: PNAD/IBGE.

Figure 1: Percentage gross enrolment rate of 18-22 year olds in tertiary education, Brazil



Source: Census (2000); PNAD (all other years).

Table 3: Multinomial logit marginal effects of years of schooling, men and women aged 15-35

agricultural employment or self- employment	other manual employment	non-manual employment	self- employed outside agriculture	in education	not in work or education
1993					
-0.0209 (-73.08)	-0.0075 (-21.37)	0.0402 (76.99)	-0.0041 (-14.18)	0.0086 (43.27)	-0.0164 (-34.81)
n = 116987		LL = -76101685		Pseudo R ² = 0.1590	
1999					
-0.0162 (-63.06)	-0.0060 (-18.82)	0.0340 (67.98)	-0.0053 (-20.30)	0.0124 (48.28)	-0.0189 (-41.55)
n = 129364		LL = -81906572		Pseudo R ² = 0.1706	
2005					
-0.0140 (-60.08)	-0.0095 (-30.43)	0.0398 (79.02)	-0.0062 (-25.14)	0.0093 (29.99)	-0.0195 (-46.34)
n = 149412		LL = -92798298		Pseudo R ² = 0.1700	
2013					
-0.0080 (-41.79)	-0.0141 (-41.40)	0.0446 (67.89)	-0.0058 (-22.28)	0.0084 (24.78)	-0.0251 (-51.14)
n = 124745		LL = -90896264		Pseudo R ² = 0.1817	

Note: z values in parentheses; control variables are as described in the text.

Table 4: Multinomial logit marginal effects of years of schooling, men aged 15-35

agricultural employment or self- employment	other manual employment	non-manual employment	self- employed outside agriculture	in education	not in work or education
1993					
-0.0299 (-57.54)	-0.0087 (-13.40)	0.0457 (61.82)	-0.0065 (-12.74)	0.0055 (20.95)	-0.0060 (-12.07)
n = 56530		LL = -37015300		Pseudo R ² =	0.1643
1999					
-0.0235 (-50.86)	-0.0060 (-9.99)	0.0380 (54.47)	-0.0084 (-18.06)	0.0097 (29.66)	-0.0098 (-18.68)
n = 62631		LL = -40849500		Pseudo R ² =	0.1732
2005					
-0.0216 (-51.53)	-0.0109 (-18.74)	0.0460 (66.37)	-0.0094 (-23.17)	0.0083 (25.56)	-0.0123 (-25.74)
n = 73196		LL = -46237241		Pseudo R ² =	0.1811
2013					
-0.0128 (-36.67)	-0.0200 (-30.30)	0.0511 (55.75)	-0.0090 (-20.63)	0.0078 (20.50)	-0.0170 (-31.84)
n = 61093		LL = -46263853		Pseudo R ² =	0.1847

Note: z values in parentheses; control variables are as described in the text.

Table 5: Multinomial logit marginal effects of years of schooling, women aged 15-35

agricultural employment or self- employment	other manual employment	non-manual employment	self- employed outside agriculture	in education	not in work or education
1993					
-0.0132 (-40.56)	-0.0018 (-7.37)	0.0348 (46.70)	0.0004 (1.20)	0.0109 (36.67)	-0.0310 (-41.99)
n = 60457		LL = -33601348		Pseudo R ² = 0.1824	
1999					
-0.0097 (-33.66)	-0.0009 (-4.68)	0.0291 (40.39)	0.0002 (0.57)	0.0135 (31.89)	-0.0322 (-45.12)
n = 66733		LL = -36226175		Pseudo R ² = 0.1902	
2005					
-0.0071 (-28.14)	-0.0022 (-10.09)	0.0321 (43.97)	-0.0004 (-1.23)	0.0077 (14.38)	-0.0301 (-44.87)
n = 76216		LL = -41629541		Pseudo R ² = 0.1754	
2013					
-0.0035 (-17.06)	-0.0025 (-12.31)	0.0353 (38.49)	-0.0001 (-0.19)	0.0069 (12.14)	-0.0362 (-44.78)
n = 63652		LL = -39736874		Pseudo R ² = 0.1950	

Note: z values in parentheses; control variables are as described in the text.

Table 6: Dynamic discrete choice model: parameter estimates

Parameter	Estimated coefficient	t statistic
α_{10}	5.0937	21120.63
α_{11}	0.0626	526.67
α_{12}	0.3483	704.22
α_{13}	0.1058	134.85
α_{20}	4.2737	1158.51
α_{21}	0.0715	420.11
α_{22}	0.4947	493.27
α_{23}	-0.0195	27.30
$\beta_0/1000$	3.2795	10.54
$\beta_1/1000$	-4.3985	13.67
$\beta_2/1000$	0.0500	2.61
$\gamma_0/1000$	-0.9451	10.48
ρ_{11}	0.0708	2001.34
ρ_{22}	0.1080	1077.96
ρ_{33}	4.8188	27.42
ρ_{44}	2.6870	26.82
Log likelihood	-360923.86	

Note: The ρ terms are the standard deviations such that:

$$\varepsilon_{1t} = \rho_{11}\eta_{1t}$$

$$\varepsilon_{2t} = \rho_{22}\eta_{2t}$$

$$\varepsilon_{3t} = \rho_{33}\eta_{3t}$$

$$\varepsilon_{4t} = \rho_{44}\eta_{4t}$$

$$\eta_{kt} \sim N(0,1), k=1,\dots,4.$$